A Lower Bound on the Noncoherent Capacity Pre-log for the MIMO Channel with Temporally Correlated Fading

Günther Koliander[1], Erwin Riegler[1], Giuseppe Durisi[2], Veniamin I. Morgenshtern[3], and Franz Hlawatsch[1]

[1]Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria
[2]Department of Signals and Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden
[3]Department of Statistics, Stanford University, CA 94305, USA

***Abstract*—We derive a lower bound on the capacity pre-log of a temporally correlated Rayleigh block-fading multiple-input multiple-output (MIMO) channel with T transmit antennas and R receive antennas in the noncoherent setting (no *a priori* channel knowledge at the transmitter and the receiver). In this model, the fading process changes independently across blocks of length L and is temporally correlated within each block for each transmit-receive antenna pair, with a given rank Q of the corresponding correlation matrix. Our result implies that for almost all choices of the coloring matrix that models the temporal correlation, the pre-log can be lower-bounded by $T(1-1/L)$ for $T \leq (L-1)/Q$ provided that R is sufficiently large. The widely used constant block-fading model is equivalent to the temporally correlated block-fading model with $Q=1$ for the special case when the temporal correlation for each transmit-receive antenna pair is the same, which is unlikely to be observed in practice. For the constant block-fading model, the capacity pre-log is given by $T(1-T/L)$, which is smaller than our lower bound for the case $Q=1$. Thus, our result suggests that the assumptions underlying the constant block-fading model lead to a pessimistic result for the capacity pre-log.**

I. Introduction

We analyze the capacity of a Rayleigh block-fading multiple-input multiple-output (MIMO) channel in the noncoherent setting where the transmitter and the receiver are aware of the channel statistics but have no *a priori* channel state information. In this setting, the penalty on capacity[1] incurred by allocating resources to channel estimation is automatically accounted for. We model channel variations in time by the *temporally correlated block-fading model* introduced in [1]. According to this model, the fading process takes on independent realizations across blocks of length L; however, for each transmit-receive antenna pair, it is correlated within each block with a given rank Q of the corresponding $L \times L$ correlation matrix.

The capacity of the temporally correlated block-fading channel is not known even in the single-input single-output (SISO) case. The capacity pre-log, which is defined as the ratio of the capacity to the logarithm of the signal-to-noise ratio (SNR) as the SNR goes to infinity, has been characterized in [1] for the SISO case and in [2]–[4] for the single-input multiple-output (SIMO) case. For *regular stationary* fading processes, the capacity of the MIMO channel has been studied in [5]. It was proved that, in this case, the capacity grows only doubly-logarithmically due to the regular fading assumption. For *nonregular stationary* fading processes, the MIMO capacity pre-log is not known to date.

In this paper, we derive a lower bound on the capacity pre-log of a rank-Q temporally correlated block-fading MIMO channel with block length L, T transmit antennas, and R receive antennas. We show that the pre-log is lower-bounded by $T(1-1/L)$ for $T \leq (L-1)/Q$ provided that $R \geq T(L-1)/(L-TQ)$. This lower bound can be achieved for almost all (a.a.) choices—i.e., up to a set of measure zero—of the coloring matrix that models the temporal correlation for the transmit-receive antenna pairs.

Our result is particularly surprising when compared to the capacity for the constant block-fading model as derived by Zheng and Tse [6]. The constant block-fading model is a special case of the temporally correlated block-fading model for $Q=1$ that is obtained when the correlation matrices for all transmit-receive antenna pairs are assumed to be equal, which is unlikely to be observed in practice. Zheng and Tse showed that the pre-log for the constant block-fading model is $M^*(1-M^*/L)$ with $M^* \triangleq \min\{T, R, \lfloor L/2 \rfloor\}$, which is less than or equal to $L/4$. In the temporally correlated block-fading model for $Q=1$, our lower bound on the pre-log is $L-2+1/L$ if $T=L-1$ and $R=(L-1)^2$ for a.a. coloring matrices.[2] This shows that a much higher pre-log can be achieved and, hence, the results predicted by the constant block-fading model are pessimistic.

Apart from our main result, the methods employed in its proof may be of independent interest. We use a generalized

This work was supported by the WWTF under grant ICT10-066 (NOWIRE).

[1]In this paper, the term *capacity* refers to capacity in the noncoherent setting.

[2]Note that the coloring matrix corresponding to the constant block-fading model belongs to the set of measure zero where this bound does not hold.

change-of-variables theorem for integrals in combination with Bézout's theorem [7, Proposition B.2.7] to establish certain transformation properties of differential entropy under finite-to-one mappings. Furthermore, we use an important property of subharmonic functions to lower-bound the integral of a certain real analytic function. In the SIMO case, a similar problem was recently solved using an algebraic-geometry method [3]. Our alternative method works in a more general setting and, thus, may also be useful for bounding differential entropy terms appearing in other problems.

The rest of this paper is organized as follows. The system model is presented in Section II. The lower bound on the capacity pre-log is stated and discussed in Section III. A proof of the lower bound is provided in Sections IV and V and in three appendices.

Notation: Sets are denoted by calligraphic letters (e.g., $\mathcal{I}$), and $|\mathcal{I}|$ denotes the cardinality of $\mathcal{I}$. Sets of sets are denoted by fraktur letters (e.g., $\mathfrak{M}$). We use the notation $[M:N] \triangleq \{M, M+1, \ldots, N\}$ for $M, N \in \mathbb{N}$. Boldface uppercase (lowercase) letters denote matrices (vectors). Sans serif letters denote random quantities, e.g., A is a random matrix and x is a random vector. The superscripts $^{\mathrm{T}}$ and $^{\mathrm{H}}$ stand for transposition and Hermitian transposition, respectively. The all-zero matrix or vector of appropriate size is written as $\mathbf{0}$, and the $M \times M$ identity matrix as $\boldsymbol{I}_M$. For a matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, the element in the ith row and jth column is denoted by $a_{i,j}$. We denote by $[\boldsymbol{A}]_{\mathcal{I}}^{\mathcal{J}}$, where $\mathcal{I} \subseteq [1:M]$ and $\mathcal{J} \subseteq [1:N]$, the $|\mathcal{I}| \times |\mathcal{J}|$ submatrix of $\boldsymbol{A}$ containing the elements $a_{i,j}$ with $i \in \mathcal{I}$ and $j \in \mathcal{J}$; furthermore, $[\boldsymbol{A}]_{\mathcal{I}} \triangleq [\boldsymbol{A}]_{\mathcal{I}}^{[1:N]}$ and $[\boldsymbol{A}]^{\mathcal{J}} \triangleq [\boldsymbol{A}]_{[1:M]}^{\mathcal{J}}$. We denote by $[\boldsymbol{x}]_{\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|}$ the subvector of $\boldsymbol{x}$ containing the elements x_i with $i \in \mathcal{I}$. The diagonal matrix with the elements of $\boldsymbol{x}$ in its main diagonal is denoted by $\mathrm{diag}(\boldsymbol{x})$. We define $\mathrm{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_K)$ as the block diagonal matrix with the matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_K$ on the main block diagonal. The modulus of the determinant of a square matrix $\boldsymbol{A}$ is denoted by $|\boldsymbol{A}|$. For $x \in \mathbb{R}$, $\lfloor x \rfloor \triangleq \max\{m \in \mathbb{Z} \mid m \le x\}$ and $\lceil x \rceil \triangleq \min\{m \in \mathbb{Z} \mid m \ge x\}$. We write $\mathbb{E}[\cdot]$ for the expectation operator, and $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the distribution of a jointly proper Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The Jacobian matrix of a differentiable function ϕ is denoted by $\boldsymbol{J}_\phi$.

II. System Model

We consider a MIMO channel with T transmit and R receive antennas. The fading process associated with each transmit-receive antenna pair conforms to the temporally correlated block-fading model [1], which results in the following channel input-output relations within a given block of length L:

$$\mathsf{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t \in [1:T]} \mathrm{diag}(\mathsf{h}_{r,t})\, \mathsf{x}_t + \mathsf{n}_r\,, \quad r \in [1\!:\!R]\,. \tag{1}$$

Here, $\mathsf{x}_t \in \mathbb{C}^L$ is the signal vector transmitted by the tth transmit antenna; $\mathsf{y}_r \in \mathbb{C}^L$ is the vector received by the rth receive antenna; $\mathsf{h}_{r,t} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{Z}_{r,t}\boldsymbol{Z}_{r,t}^{\mathrm{H}})$, where $\boldsymbol{Z}_{r,t} \in \mathbb{C}^{L \times Q}$ with $Q \triangleq \mathrm{rank}(\boldsymbol{Z}_{r,t}\boldsymbol{Z}_{r,t}^{\mathrm{H}})$, is the vector of channel coefficients between the tth transmit antenna and the rth receive antenna; $\mathsf{n}_r \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_L)$ is the noise vector at the rth receive antenna; and $\rho \in \mathbb{R}^+$ is the SNR. The vectors $\mathsf{h}_{r,t}$ and n_r are assumed to be mutually independent and independent across $r \in [1\!:\!R]$ and $t \in [1\!:\!T]$, and to change in an independent fashion from block to block ("block-memoryless" assumption). The transmitted signal vectors x_t are assumed to be independent of the vectors $\mathsf{h}_{r,t}$ and n_r. We note that the channel coefficient vectors can be written as

$$\mathsf{h}_{r,t} = \boldsymbol{Z}_{r,t}\mathsf{s}_{r,t}\,,$$

with the Q-dimensional whitened vectors $\mathsf{s}_{r,t} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_Q)$.

Setting $\mathsf{y} \triangleq (\mathsf{y}_1^{\mathrm{T}}, \ldots, \mathsf{y}_R^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RL}$ and $\mathsf{n} \triangleq (\mathsf{n}_1^{\mathrm{T}}, \ldots, \mathsf{n}_R^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RL}$, the R input-output relations (1) can be written more compactly as

$$\mathsf{y} = \sqrt{\frac{\rho}{T}}\, \bar{\mathsf{y}} + \mathsf{n}\,, \tag{2}$$

with

$$\bar{\mathsf{y}} \triangleq \sum_{t \in [1:T]} \begin{pmatrix} \mathsf{X}_t \boldsymbol{Z}_{1,t} & & \\ & \ddots & \\ & & \mathsf{X}_t \boldsymbol{Z}_{R,t} \end{pmatrix} \mathsf{s}_t \in \mathbb{C}^{RL}\,, \tag{3}$$

where we have defined $\mathsf{X}_t \triangleq \mathrm{diag}(\mathsf{x}_t) \in \mathbb{C}^{L \times L}$ and $\mathsf{s}_t \triangleq (\mathsf{s}_{1,t}^{\mathrm{T}}, \ldots, \mathsf{s}_{R,t}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RQ}$. For later use, we also define $\mathsf{x} \triangleq (\mathsf{x}_1^{\mathrm{T}}, \ldots, \mathsf{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{TL}$, $\mathsf{s} \triangleq (\mathsf{s}_1^{\mathrm{T}}, \ldots, \mathsf{s}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{TRQ}$, and

$$\boldsymbol{Z} \triangleq \begin{pmatrix} \boldsymbol{Z}_{1,1} & \cdots & \boldsymbol{Z}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{Z}_{R,1} & \cdots & \boldsymbol{Z}_{R,T} \end{pmatrix} \in \mathbb{C}^{RL \times TQ}.$$

We will refer to $\boldsymbol{Z}$ as the *coloring matrix*.

III. A Lower Bound on the Capacity Pre-log

Because of the block-memoryless assumption, the coding theorem in [8, Section 7.3] implies that the capacity of the channel (2) is given by

$$C(\rho) = \frac{1}{L} \sup_f I(\mathsf{x}; \mathsf{y})\,. \tag{4}$$

Here, $I(\mathsf{x}; \mathsf{y})$ denotes mutual information [9, p. 251] and the supremum is taken over all input distributions f on $\mathbb{C}^{TL}$ that satisfy the average power constraint

$$\mathbb{E}[\|\mathsf{x}\|^2] \le TL\,.$$

The capacity pre-log is then defined as

$$\chi \triangleq \lim_{\rho \to \infty} \frac{C(\rho)}{\log(\rho)}\,. \tag{5}$$

We will obtain the main result of this paper, which is stated in Theorem 1 below, by maximizing with respect to T the lower bound on χ given in the following proposition.

Proposition 1: For $T \le R$, there exists a set $\mathcal{Z} \subseteq \mathbb{C}^{RL \times TQ}$ with a complement of Lebesgue measure zero such that for each coloring matrix $\boldsymbol{Z} \in \mathcal{Z}$, the capacity pre-log of the channel (1) satisfies

$$\chi \ \geq\ \chi_{\mathrm{low}}(T) \ \triangleq\ \min\left\{T\left(1-\frac{1}{L}\right), R\left(1-\frac{TQ}{L}\right)\right\}. \quad (6)$$

Proof: See Section IV. ∎

The main result of this paper is stated in the following theorem.

Theorem 1: There exists a set $\mathcal{Z} \subseteq \mathbb{C}^{RL\times TQ}$ with a complement of Lebesgue measure zero such that for each coloring matrix $\boldsymbol{Z} \in \mathcal{Z}$, the capacity pre-log of the channel (1) satisfies

$$\chi \ \geq\ \chi^{*}_{\mathrm{low}} \ \triangleq\ \begin{cases} T\left(1-\dfrac{1}{L}\right) & \text{if } T \leq T_{\mathrm{opt}} \\ \eta & \text{if } T > T_{\mathrm{opt}}\,, \end{cases} \quad (7)$$

where

$$\eta \ \triangleq\ \max\left\{R\left(1-\frac{\lceil T_{\mathrm{opt}}\rceil Q}{L}\right), \lfloor T_{\mathrm{opt}}\rfloor\left(1-\frac{1}{L}\right)\right\} \quad (8)$$

and

$$T_{\mathrm{opt}} \ \triangleq\ \frac{RL}{L+RQ-1} \ \leq\ \min\{R, L/Q\}\,. \quad (9)$$

Proof: We obtain a lower bound on the pre-log for T transmit antennas by maximizing χ_{low} with respect to the number of effectively used transmit antennas (note that we can always switch off some antennas). Thus, we will take the maximum of $\chi_{\mathrm{low}}(T^{*})$ in (6) with respect to $T^{*} \leq \min\{T, R\}$. Here, T^{*} is also restricted by $T^{*} \leq R$ because Proposition 1 holds only for $T^{*} \leq R$. Because $\chi_{\mathrm{low}}(T^{*})$ is the minimum of two quantities where the first is monotonically increasing in T^{*} and the second is monotonically decreasing in T^{*}, it attains its maximum at the intersection point T_{opt} defined in (9). If $T \leq T_{\mathrm{opt}}$, we have $\chi_{\mathrm{low}}(T) = T(1-1/L)$, which proves the first case in (7). For $T > T_{\mathrm{opt}}$, we have to take into account that T_{opt} might not be a natural number. Thus, we have to take the maximum of $\chi_{\mathrm{low}}(\lfloor T_{\mathrm{opt}}\rfloor)$ and $\chi_{\mathrm{low}}(\lceil T_{\mathrm{opt}}\rceil)$, which turns out to be η in (8). This shows the second case in (7) and concludes the proof. ∎

Remark 1: The set $\mathcal{Z}$ will be specified in Definition 1 in Section V.

Remark 2: For a fixed R, the maximum value of χ^{*}_{low} in (7) is obtained by using either $\lfloor T_{\mathrm{opt}}\rfloor$ or $\lceil T_{\mathrm{opt}}\rceil$ transmit antennas. This implies that the optimal number of transmit antennas is upper-bounded by $\lceil T_{\mathrm{opt}}\rceil$.

Remark 3: For $L = Q$, χ^{*}_{low} is equal to zero and hence trivial.

Remark 4: The lower bound χ^{*}_{low} in (7) can be expressed as

$$\chi^{*}_{\mathrm{low}} \ =\ \min\left\{T\left(1-\frac{1}{L}\right), \eta\right\}.$$

Remark 5: χ^{*}_{low} can be at most $\lfloor (L-1)/Q\rfloor(1-1/L)$. This value of χ^{*}_{low} is attained for $T = \lfloor (L-1)/Q\rfloor$ and $R = \lceil (L-1)^2/Q\rceil$.

Remark 6: By (9), the condition $T \leq T_{\mathrm{opt}}$ in (7) is equivalent to $R \geq T(L-1)/(L-TQ)$. Thus, for a fixed $T < L/Q$, we can always obtain $\chi^{*}_{\mathrm{low}} = T(1-1/L)$ by using a sufficiently large R.

Remark 7: If all matrices $\boldsymbol{Z}_{r,t}$ for $r \in [1\!:\!R]$ and $t \in [1\!:\!T]$ coincide, the temporally correlated block-fading model for $Q = 1$ reduces to the constant block-fading model studied in [6]. The pre-log in the constant block-fading model is $M^{*}(1-M^{*}/L)$, where $M^{*} \triangleq \min\{T, R, \lfloor L/2\rfloor\}$; therefore, it is upper-bounded by $L/4$. On the other hand, Theorem 1 for $Q = 1$ implies that the pre-log for the correlated block-fading model is lower-bounded by (cf. (6))

$$\chi \geq \min\left\{T\left(1-\frac{1}{L}\right), R\left(1-\frac{T}{L}\right)\right\}, \quad (10)$$

for a.a. coloring matrices $\boldsymbol{Z} \in \mathcal{Z}$. In particular, for $T = L-1$ and $R = (L-1)^2$, the lower bound in (10) becomes $L-2+1/L$. Thus, for a.a. choices of coloring matrices, the pre-log is much higher than in the constant block-fading model.[3] Hence, the results predicted by the constant block-fading model are pessimistic.

IV. Proof of Proposition 1

For $L \leq TQ$, the inequality in (6) is trivially true, because in this case $\chi_{\mathrm{low}} \leq 0$. Therefore, it remains to prove (6) for $L > TQ$, which will thus be assumed in the following. By (4), the capacity can be lower-bounded as $C(\rho) \geq (1/L)I(\mathbf{x};\mathbf{y})$ with the specific input distribution $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{TL})$. Inserting this lower bound into (5) then gives

$$\chi \ \geq\ \frac{1}{L}\lim_{\rho\to\infty}\frac{I(\mathbf{x};\mathbf{y})}{\log(\rho)}\,. \quad (11)$$

In what follows, we thus assume that $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{TL})$.

We have $I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}\,|\,\mathbf{x})$ with h denoting differential entropy. Hence, we can lower-bound $I(\mathbf{x};\mathbf{y})$ by upper-bounding $h(\mathbf{y}\,|\,\mathbf{x})$ and lower-bounding $h(\mathbf{y})$. Similar to [2, Eq. (8)], we have

$$h(\mathbf{y}\,|\,\mathbf{x}) \ \leq\ TQR\log(\rho) + \mathcal{O}(1)\,, \quad (12)$$

where "$+\,\mathcal{O}(1)$" means "up to a function of ρ that is bounded for $\rho \to \infty$." Furthermore, similar to [2, Eq. (12)], we have

$$h(\mathbf{y}) \geq \left(\sum_{r\in[1:R]} |\mathcal{I}_r|\right)\log(\rho) + h(\boldsymbol{P}\bar{\mathbf{y}}) + c\,, \quad (13)$$

where $\bar{\mathbf{y}}$ was defined in (3),

$$\boldsymbol{P} \ \triangleq\ \mathrm{diag}\big([\boldsymbol{I}_L]_{\mathcal{I}_1}, \ldots, [\boldsymbol{I}_L]_{\mathcal{I}_R}\big) \in \mathbb{C}^{\sum_{r\in[1:R]}|\mathcal{I}_r|\times RL}, \quad (14)$$

the $\mathcal{I}_r \subseteq [1\!:\!L]$ for $r \in [1\!:\!R]$ are certain subsets that will be specified later, and c is a finite constant. Note that in (13), $h(\boldsymbol{P}\bar{\mathbf{y}})$ and c do not depend on ρ. Using (12) and (13) in $I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}\,|\,\mathbf{x})$, we obtain

$$I(\mathbf{x};\mathbf{y}) \geq \left(\sum_{r\in[1:R]} |\mathcal{I}_r| - TQR\right)\log(\rho) + h(\boldsymbol{P}\bar{\mathbf{y}}) + \mathcal{O}(1)\,. \quad (15)$$

The proposed lower bound on the pre-log in (6) is established by inserting (15) into (11) and choosing the sets $\{\mathcal{I}_r\}_{r\in[1:R]}$

[3] This implies that the coloring matrices corresponding to the constant block-fading model belong to the complement of $\mathcal{Z}$ (which has Lebesgue measure zero and is unlikely to be observed in practice).

such that

$$\sum_{r\in[1:R]} |\mathcal{I}_r| \;=\; \min\{TL-T+TQR, RL\}\,, \tag{16}$$

provided that $h(\boldsymbol{P}\bar{\mathbf{y}}) > -\infty$. It remains to show that there exist sets $\{\mathcal{I}_r\}_{r\in[1:R]}$ satisfying (16) and a set $\mathcal{Z} \subseteq \mathbb{C}^{RL\times TQ}$ with a complement of Lebesgue measure zero for which $h(\boldsymbol{P}\bar{\mathbf{y}}) > -\infty$ for each $\boldsymbol{Z} \in \mathcal{Z}$. This will be done in the next section.

V. Proof that $h(\boldsymbol{P}\bar{\mathbf{y}}) > -\infty$

Let us split the vector $\mathbf{x}$ into the vectors $\mathbf{x}_{\mathcal{P}} \triangleq \big([\mathbf{x}_1]^{\mathrm{T}}_{\mathcal{P}_1}, \dots, [\mathbf{x}_T]^{\mathrm{T}}_{\mathcal{P}_T}\big)^{\mathrm{T}}$ and $\mathbf{x}_{\mathcal{D}} \triangleq \big([\mathbf{x}_1]^{\mathrm{T}}_{\mathcal{D}_1}, \dots, [\mathbf{x}_T]^{\mathrm{T}}_{\mathcal{D}_T}\big)^{\mathrm{T}}$, where $\mathcal{P}_t \subseteq [1\!:\!L]$ and $\mathcal{D}_t \triangleq [1\!:\!L]\backslash\mathcal{P}_t$ for $t\in[1\!:\!T]$. Because $h(\boldsymbol{P}\bar{\mathbf{y}}|\mathbf{x}_{\mathcal{P}}) \le h(\boldsymbol{P}\bar{\mathbf{y}})$, it is sufficient to show that $h(\boldsymbol{P}\bar{\mathbf{y}}|\mathbf{x}_{\mathcal{P}}) > -\infty$. As in [3], we wish to relate $h(\boldsymbol{P}\bar{\mathbf{y}}|\mathbf{x}_{\mathcal{P}})$ to the simpler quantity $h(\mathbf{s},\mathbf{x}_{\mathcal{D}}) = h(\mathbf{s}) + h(\mathbf{x}_{\mathcal{D}})$. This will be done via the family of $\mathbf{x}_{\mathcal{P}}$-parametrized mappings

$$\phi_{\boldsymbol{x}_{\mathcal{P}}} : (\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \mapsto \boldsymbol{P}\bar{\boldsymbol{y}}\,, \tag{17}$$

where $\bar{\boldsymbol{y}}$ is defined in (3), i.e.,

$$\bar{\boldsymbol{y}} \;=\; \sum_{t\in[1:T]} \boldsymbol{\Xi}_t \boldsymbol{s}_t\,, \tag{18}$$

with

$$\boldsymbol{\Xi}_t \;\triangleq\; \begin{pmatrix} \boldsymbol{X}_t\boldsymbol{Z}_{1,t} & & \\ & \ddots & \\ & & \boldsymbol{X}_t\boldsymbol{Z}_{R,t}\end{pmatrix} \in \mathbb{C}^{RL\times RQ}. \tag{19}$$

According to (18) and (19), the components of each vector-valued mapping $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ are multivariate polynomials of degree 2. The Jacobian matrix $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}$ of each mapping $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ is equal to

$$\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s},\boldsymbol{x}_{\mathcal{D}}) \;=\; \boldsymbol{P}\begin{pmatrix} \boldsymbol{\Xi}_1 \cdots \boldsymbol{\Xi}_T & \begin{matrix}\boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,T}\\ \vdots & & \vdots \\ \boldsymbol{A}_{R,1} & \cdots & \boldsymbol{A}_{R,T}\end{matrix}\end{pmatrix} \in \mathbb{C}^{\sum_{r\in[1:R]}|\mathcal{I}_r|\times\left(TQR+\sum_{t\in[1:T]}|\mathcal{D}_t|\right)}, \tag{20}$$

where

$$\boldsymbol{A}_{r,t} \;\triangleq\; \left[\operatorname{diag}\left(a^{(1)}_{r,t},\dots,a^{(L)}_{r,t}\right)\right]^{\mathcal{D}_t}, \quad t\in[1\!:\!T],\, r\in[1\!:\!R]\,,$$
$$\text{with}\;\; a^{(\ell)}_{r,t} \triangleq [\boldsymbol{Z}_{r,t}]_{\{\ell\}}\boldsymbol{s}_{r,t}\,,\quad \ell\in[1\!:\!L]\,. \tag{21}$$

Note that by (14), $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s},\boldsymbol{x}_{\mathcal{D}})$ can be written as

$$\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s},\boldsymbol{x}_{\mathcal{D}}) \;=\; \begin{pmatrix} \tilde{\boldsymbol{\Xi}}_1 \cdots \tilde{\boldsymbol{\Xi}}_T & \begin{matrix}[\boldsymbol{A}_{1,1}]_{\mathcal{I}_1} & \cdots & [\boldsymbol{A}_{1,T}]_{\mathcal{I}_1}\\ \vdots & & \vdots \\ [\boldsymbol{A}_{R,1}]_{\mathcal{I}_R} & \cdots & [\boldsymbol{A}_{R,T}]_{\mathcal{I}_R}\end{matrix}\end{pmatrix}, \tag{22}$$

where

$$\tilde{\boldsymbol{\Xi}}_t \;\triangleq\; \begin{pmatrix} [\boldsymbol{X}_t\boldsymbol{Z}_{1,t}]_{\mathcal{I}_1} & & \\ & \ddots & \\ & & [\boldsymbol{X}_t\boldsymbol{Z}_{R,t}]_{\mathcal{I}_R}\end{pmatrix}.$$

Based on the family of mappings $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ in (17), the relation between $h(\boldsymbol{P}\bar{\mathbf{y}}|\mathbf{x}_{\mathcal{P}})$ and $h(\mathbf{s},\mathbf{x}_{\mathcal{D}})$ can be established by using the definition of conditional differential entropy [9, Chapter 8] and by applying the change-of-variables theorem for integrals under finite-to-one mappings[4] [10, Theorem 3.2.5]. For this, we need to show that the family of mappings $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ is finite-to-one almost everywhere (a.e.) for a.a. choices of $\boldsymbol{x}_{\mathcal{P}}$. We now define the set $\mathcal{Z}$ for which this proof works.

Definition 1: Let $\mathcal{Z} \subseteq \mathbb{C}^{RL\times TQ}$ be the set of matrices $\boldsymbol{Z}$ such that the following holds: There exist a choice of sets $\{\mathcal{I}_r\}_{r\in[1:R]}$ satisfying (16), i.e.,

$$\sum_{r\in[1:R]} |\mathcal{I}_r| \;=\; \min\{TL-T+TQR, RL\}\,, \tag{23}$$

and a choice of sets $\{\mathcal{P}_t\}_{t\in[1:T]}$ satisfying

$$\sum_{t\in[1:T]} |\mathcal{P}_t| \;=\; \max\{T, TQR-(R-T)L\}\,, \tag{24}$$

such that $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s},\boldsymbol{x}_{\mathcal{D}})$ is nonsingular a.e. for a.a. choices of $\boldsymbol{x}_{\mathcal{P}}$.

We will show presently that the set $\mathcal{Z}$ is nonempty. In fact, it covers a.a. of $\mathbb{C}^{RL\times TQ}$.

Condition (23) on $\{|\mathcal{I}_r|\}_{r\in[1:R]}$ and condition (24) on $\{|\mathcal{P}_t|\}_{t\in[1:T]}$ guarantee that the matrix $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s},\boldsymbol{x}_{\mathcal{D}})$ is square. More specifically, we have with (20) that

$$\#\text{rows} \;=\; \sum_{r\in[1:R]} |\mathcal{I}_r| \;=\; \min\{TL-T+TQR, RL\}\,, \tag{25}$$

where (23) was used, and

$$\begin{aligned}\#\text{columns} \;&=\; TQR + \sum_{t\in[1:T]} |\mathcal{D}_t| &(26)\\ &=\; TQR + TL - \sum_{t\in[1:T]} |\mathcal{P}_t| \\ &=\; TQR + TL - \max\{T, TQR-(R-T)L\} \\ &=\; \min\{TQR+TL-T, RL\}\,, &(27)\end{aligned}$$

where (24) was used. Thus, comparing (25) and (27), we have $\#\text{rows} = \#\text{columns}$.

The next lemma states that $\mathcal{Z}$ satisfies one of the claims made in Proposition 1.

Lemma 1: The complement of the set $\mathcal{Z}$ has Lebesgue measure zero.

Proof: See Appendix A. ∎

In the remainder of our proof that $h(\boldsymbol{P}\bar{\mathbf{y}}) > -\infty$, we consider an arbitrary $\boldsymbol{Z}\in\mathcal{Z}$. To use the change-of-variables theorem, we will invoke Bézout's theorem to show that the mappings $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ are finite-to-one a.e.

Lemma 2: Let $\tilde{\mathcal{M}}$ be defined as the set of all $(\boldsymbol{s},\boldsymbol{x}_{\mathcal{D}})$ such that $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s},\boldsymbol{x}_{\mathcal{D}})$ is nonsingular. Then for all $\boldsymbol{y}\in\phi_{\boldsymbol{x}_{\mathcal{P}}}(\tilde{\mathcal{M}})$,

[4] For a finite-to-one mapping, the inverse image of each point in the codomain is a set of finite cardinality.

we have

$$|\phi_{\boldsymbol{x}_{\mathcal{P}}}^{-1}(\{\boldsymbol{y}\}) \cap \tilde{\mathcal{M}}| \leq \tilde{m} \triangleq 2^{\left(\sum_{t\in[1:T]}|\mathcal{D}_t| + TQR\right)}. \quad (28)$$

Proof: Let $\boldsymbol{y} \in \phi_{\boldsymbol{x}_{\mathcal{P}}}(\tilde{\mathcal{M}})$. Then according to (17)–(19), the zeros of the vector-valued mapping

$$(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \mapsto \phi_{\boldsymbol{x}_{\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) - \boldsymbol{y}$$

are the common zeros of $\sum_{t\in[1:T]}|\mathcal{D}_t| + TQR$ polynomials of degree 2. Thus, by a weak version of Bézout's theorem [7, Proposition B.2.7], the number of isolated zeros (i.e., with no other zeros in some neighborhood) cannot exceed $\tilde{m}$. Since $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}$ is nonsingular on $\tilde{\mathcal{M}}$, the function $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ restricted to $\tilde{\mathcal{M}}$ is locally one-to-one and, hence, $\phi_{\boldsymbol{x}_{\mathcal{P}}} - \boldsymbol{y}$ has only isolated zeros on $\tilde{\mathcal{M}}$. Therefore, the number of points $(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \in \tilde{\mathcal{M}}$ such that $\phi_{\boldsymbol{x}_{\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) = \boldsymbol{y}$ cannot exceed $\tilde{m}$. ■

Next, we will establish a transformation property of differential entropy under finite-to-one mappings in a general setting. More specifically, we will obtain a lower bound on differential entropy using the change-of-variables theorem for finite-to-one mappings [10, Theorem 3.2.5] in combination with the uniform bound in Lemma 2.

Lemma 3: Let $\mathsf{u} \in \mathbb{C}^n$ be a random vector with continuous density function f_{u}. Furthermore, let $\kappa\colon \mathbb{C}^n \to \mathbb{C}^n$ be a continuously differentiable mapping with Jacobian matrix $\boldsymbol{J}_\kappa$ and let $\mathcal{M} \triangleq \{\boldsymbol{u} \in \mathbb{C}^n : |\boldsymbol{J}_\kappa(\boldsymbol{u})| \neq 0\}$ and $\mathsf{v} \triangleq \kappa(\mathsf{u})$. Assume that the complement of $\mathcal{M}$ has Lebesgue measure zero and $|\kappa^{-1}(\{\boldsymbol{v}\}) \cap \mathcal{M}| \leq m < \infty$ for all $\boldsymbol{v} \in \mathbb{C}^n$, with some constant $m \in \mathbb{N}$. Then there exists a set $\mathcal{U} \subseteq \mathbb{C}^n$ such that

$$h(\mathsf{v}) \geq -m\log(m) - m\int_{\mathcal{U}} f_{\mathsf{u}}(\boldsymbol{u})\log(f_{\mathsf{u}}(\boldsymbol{u}))\,d\boldsymbol{u} + m\int_{\mathcal{U}} f_{\mathsf{u}}(\boldsymbol{u})\log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\,d\boldsymbol{u}.$$

Proof: See Appendix B. ■

To lower-bound $h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}})$, we first lower-bound the differential entropies $h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}} = \boldsymbol{x}_{\mathcal{P}})$. By Lemma 2, we have $|\phi_{\boldsymbol{x}_{\mathcal{P}}}^{-1}(\{\boldsymbol{y}\}) \cap \tilde{\mathcal{M}}| \leq \tilde{m}$. Furthermore, since we assume $\boldsymbol{Z} \in \mathcal{Z}$, we have by Definition 1 that $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})$ is nonsingular a.e. and, hence, the complement of $\tilde{\mathcal{M}}$ has Lebesgue measure zero. Thus, we can invoke Lemma 3 with $h(\mathsf{v}) = h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}} = \boldsymbol{x}_{\mathcal{P}})$, $\kappa = \phi_{\boldsymbol{x}_{\mathcal{P}}}$, $\mathsf{u} = (\mathsf{s}, \mathsf{x}_{\mathcal{D}})$, and $m = \tilde{m}$ to obtain

$$\begin{aligned} h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}} = \boldsymbol{x}_{\mathcal{P}}) &\geq -\tilde{m}\log(\tilde{m}) \\ &- \tilde{m}\int_{\mathcal{U}} f_{\mathsf{s},\mathsf{x}_{\mathcal{D}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})\log(f_{\mathsf{s},\mathsf{x}_{\mathcal{D}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}))\,d(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \\ &+ \tilde{m}\int_{\mathcal{U}} f_{\mathsf{s},\mathsf{x}_{\mathcal{D}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})\log(|\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})|^2)\,d(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}). \end{aligned} \quad (29)$$

Using (29), we can now lower-bound $h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}})$ as follows:

$$\begin{aligned} h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}}) &= \int f_{\mathsf{x}_{\mathcal{P}}}(\boldsymbol{x}_{\mathcal{P}})\, h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}} = \boldsymbol{x}_{\mathcal{P}})\,d\boldsymbol{x}_{\mathcal{P}} \\ &\geq \int f_{\mathsf{x}_{\mathcal{P}}}(\boldsymbol{x}_{\mathcal{P}}) \Big[-\tilde{m}\log(\tilde{m}) \\ &\quad - \tilde{m}\int_{\mathcal{U}} f_{\mathsf{s},\mathsf{x}_{\mathcal{D}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})\log(f_{\mathsf{s},\mathsf{x}_{\mathcal{D}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}))\,d(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \\ &\quad + \tilde{m}\int_{\mathcal{U}} f_{\mathsf{s},\mathsf{x}_{\mathcal{D}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})\log(|\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})|^2)\,d(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \Big] d\boldsymbol{x}_{\mathcal{P}}. \end{aligned} \quad (30)$$

The lower bound in (30) consists of three terms. The first term is just a finite constant. The second term is finite because the differential entropy of the Gaussian random vector $(\mathsf{s}, \mathsf{x}_{\mathcal{D}})$ is finite. The last term is finite if

$$\int_{\mathbb{C}^{TL+TQR}} f_{\mathsf{s},\mathsf{x}}(\boldsymbol{s}, \boldsymbol{x})\log(|\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})|^2)\,d(\boldsymbol{s}, \boldsymbol{x}) \quad (31)$$

is finite. To show that (31) is finite, we will invoke the following general result for analytic functions.

Lemma 4: Let f be an analytic function on $\mathbb{C}^N$ that is not identically zero. Then

$$I_1 \triangleq \int_{\mathbb{C}^N} \exp(-\|\boldsymbol{\xi}\|^2)\log(|f(\boldsymbol{\xi})|)\,d\boldsymbol{\xi} > -\infty. \quad (32)$$

Proof: See Appendix C. ■

Since $f_{\mathsf{s},\mathsf{x}}$ is the density of a standard multivariate Gaussian random vector and $\det(\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}))$ is a complex polynomial that is not identically zero due to the definition of $\mathcal{Z}$ in Definition 1, the integral in (31) is finite by Lemma 4. Hence, with (30), we obtain $h(\boldsymbol{P}\bar{\mathsf{y}}|\mathsf{x}_{\mathcal{P}}) > -\infty$. This concludes the proof that $h(\boldsymbol{P}\bar{\mathsf{y}}) > -\infty$.

Appendix A: Proof of Lemma 1

We can view $\det(\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}))$ as a function $f(\boldsymbol{Z}, \boldsymbol{x}, \boldsymbol{s})$. Assume that there is a choice of sets $\{\mathcal{I}_r\}_{r\in[1:R]}$ satisfying (23) and a choice of $\{\mathcal{P}_t\}_{t\in[1:T]}$ satisfying (24) such that $f(\boldsymbol{Z}_0, \boldsymbol{x}_0, \boldsymbol{s}_0) \neq 0$ at some $(\boldsymbol{Z}_0, \boldsymbol{x}_0, \boldsymbol{s}_0)$. Thus, because for fixed $\boldsymbol{x}_0$ and $\boldsymbol{s}_0$ the function $f(\boldsymbol{Z}, \boldsymbol{x}_0, \boldsymbol{s}_0)$ is a polynomial in the entries of $\boldsymbol{Z}$ and hence analytic in $\boldsymbol{Z}$, there is a set $\tilde{\mathcal{Z}} \subseteq \mathbb{C}^{RL\times TQ}$ with a complement of Lebesgue measure zero such that $f(\boldsymbol{Z}, \boldsymbol{x}_0, \boldsymbol{s}_0) \neq 0$ for all $\boldsymbol{Z} \in \tilde{\mathcal{Z}}$. Hence, for each fixed $\boldsymbol{Z}_1 \in \tilde{\mathcal{Z}}$, $f(\boldsymbol{Z}_1, \boldsymbol{x}, \boldsymbol{s})$ is not identically zero; furthermore, it is analytic in $\boldsymbol{x}$ and $\boldsymbol{s}$. Therefore, it is nonzero for a.a. $(\boldsymbol{x}, \boldsymbol{s})$. We conclude that $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})$ is nonsingular and thus $\boldsymbol{Z}_1 \in \mathcal{Z}$. Definition 1 implies that $\tilde{\mathcal{Z}} \subseteq \mathcal{Z}$, and hence the complement of $\mathcal{Z}$ has Lebesgue measure zero.

It remains to find choices of $\{\mathcal{I}_r\}_{r\in[1:R]}$ and $\{\mathcal{P}_t\}_{t\in[1:T]}$ such that $f(\boldsymbol{Z}, \boldsymbol{x}, \boldsymbol{s}) \neq 0$ at some $(\boldsymbol{Z}, \boldsymbol{x}, \boldsymbol{s})$. We start by choosing sets $\{\mathcal{I}_r\}_{r\in[1:R]}$ that satisfy (23). Let $k \triangleq \min\{\lfloor (TL-T)/(L-TQ)\rfloor, R\}$ and $\ell \triangleq TL - T - (L-TQ)\lfloor (TL-T)/(L-TQ)\rfloor$, and define

$$\mathcal{I}_r \triangleq \begin{cases} [1:L], & \text{if } r \in [1:k] \\ [1:TQ+\ell], & \text{if } r = k+1 \\ [1:TQ], & \text{if } r \in [k+2:R]. \end{cases} \quad (33)$$

For this choice, $[1:TQ] \subseteq \mathcal{I}_r$ for all $r \in [1:R]$, and as many $\mathcal{I}_r$ as possible without violating (23) are equal to $[1:L]$. The sets $\{\mathcal{P}_t\}_{t\in[1:T]}$ have to satisfy (cf. (24))

$$\sum_{t\in[1:T]} |\mathcal{P}_t| = \max\{T, TQR - (R-T)L\} \triangleq \vartheta_R. \quad (34)$$

We define the sets $\mathcal{P}_t$ such that $1 \in \mathcal{P}_1$, $2 \in \mathcal{P}_2, \dots, T \in \mathcal{P}_T$, and further $T+1 \in \mathcal{P}_1$, $T+2 \in \mathcal{P}_2$, etc., up to $L \in \mathcal{P}_{L \bmod T}$.

If (34) is not yet satisfied, we look for the minimal t' such that $|\mathcal{P}_{t'}|$ is minimal and $1 \notin \mathcal{P}_{t'}$ and start again with $1 \in \mathcal{P}_{t'}$, $2 \in \mathcal{P}_{t'+1}, \ldots$ We proceed until (34) is satisfied. This construction of the sets $\mathcal{P}_t$ can be formulated as

$$\mathcal{P}_t \triangleq \Big\{ i \in [1\!:\!L] : \exists j \in [1\!:\!\vartheta_R] \text{ such that } i \equiv j \bmod L \text{ and } j + \left\lfloor \frac{j-1}{\operatorname{lcm}(T,L)} \right\rfloor \equiv t \bmod T \Big\}, \quad (35)$$

where $\operatorname{lcm}(\cdot,\cdot)$ denotes the least common multiple. For example, for $T = R = 3$, $L = 6$, and $Q = 1$, we have $\vartheta_R = 9$ and (35) yields $\mathcal{P}_1 = \{1,4,3\}$, $\mathcal{P}_2 = \{2,5,1\}$, and $\mathcal{P}_3 = \{3,6,2\}$. Note that since the sizes of the sets $\mathcal{P}_t$ differ at most by 1, (35) together with (34) yields

$$\begin{aligned} |\mathcal{P}_t| &\le \left\lceil \frac{\max\{T, TQR-(R-T)L\}}{T} \right\rceil \\ &\le \left\lceil \frac{\max\{T, TQR-(R-T)TQ\}}{T} \right\rceil \\ &= TQ\,, \end{aligned} \quad (36)$$

where $L > TQ$ has been used. Some properties of the sets $\mathcal{P}_t$ are summarized in the following lemma, whose proof is omitted due to space limitations.

Lemma 5: Suppose that $R > T$. Let $\tilde{\mathcal{P}}_t \in [1\!:\!L]$ be defined according to (35) but for $R-1$ receive antennas (i.e., R is formally replaced by $R-1$) and set $\mathcal{L}_t \triangleq \tilde{\mathcal{P}}_t \backslash \mathcal{P}_t$. Then

(i) $\mathcal{L}_t \cap \mathcal{L}_{t'} = \emptyset$ for $t \neq t'$

(ii) $\mathcal{L}_t \subseteq \mathcal{I}_R$

(iii) There exist pairwise disjoint sets $\mathcal{G}_t$ satisfying $|\mathcal{G}_t| = Q$, $\mathcal{G}_t \cap \mathcal{P}_t \neq \emptyset$, and $\mathcal{G} \triangleq \bigcup_{t\in[1:T]} \mathcal{G}_t = \mathcal{I}_R \backslash \bigcup_{t\in[1:T]} \mathcal{L}_t$.

We will also make repeated use of the following result, which is a corollary of [11, pp. 21–22].

Lemma 6: Let $\boldsymbol{M} \in \mathbb{C}^{N\times N}$, and let $\mathcal{I}, \mathcal{J} \subseteq [1:N]$ with $|\mathcal{I}| = |\mathcal{J}|$. If $[\boldsymbol{M}]^{\mathcal{J}}_{[1:N]\backslash\mathcal{I}} = \boldsymbol{0}$ or $[\boldsymbol{M}]^{[1:N]\backslash\mathcal{J}}_{\mathcal{I}} = \boldsymbol{0}$, and if $[\boldsymbol{M}]^{\mathcal{J}}_{\mathcal{I}}$ is nonsingular, then $\det(\boldsymbol{M}) \neq 0$ if and only if $\det\Big([\boldsymbol{M}]^{[1:N]\backslash\mathcal{J}}_{[1:N]\backslash\mathcal{I}}\Big) \neq 0$.

Remark 8: Lemma 6 is just an abstract way to describe a situation where given a matrix $\boldsymbol{M}$, one is able to make row and column interchanges that yield a new matrix of the form $\left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{C} \end{smallmatrix}\right)$ where $\boldsymbol{A}$ and $\boldsymbol{C}$ are square matrices. In this case, it is a basic result that the determinant of $\boldsymbol{M}$ equals the product of the determinants of $\boldsymbol{A}$ and $\boldsymbol{C}$.

For the choices of $\{\mathcal{P}_t\}_{t\in[1:T]}$ and $\{\mathcal{I}_r\}_{r\in[1:R]}$ described above, it now remains to find $\boldsymbol{x}$, $\boldsymbol{s}$, and $\boldsymbol{Z}$ such that $f(\boldsymbol{Z},\boldsymbol{x},\boldsymbol{s}) = \det(\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s},\boldsymbol{x}_\mathcal{D}))$ is nonzero. This will be done by an induction argument over $R \ge T$.

Induction hypothesis: For $R \ge T$ (as assumed in Proposition 1), $\{\mathcal{P}_t\}_{t\in[1:T]}$ as in (35), and $\{\mathcal{I}_r\}_{r\in[1:R]}$ as in (33), there exists a point $(\boldsymbol{Z},\boldsymbol{x},\boldsymbol{s})$ with $\boldsymbol{x} = (1,\ldots,1)^{\mathrm{T}}$ such that $f(\boldsymbol{Z},\boldsymbol{x},\boldsymbol{s}) = \det(\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s},\boldsymbol{x}_\mathcal{D}))$ is nonzero.

Base case (proof for $R = T$): We have to show that the determinant of the matrix in (22) is nonzero for $R = T$. For $R = T$, (34) reduces to $\sum_{t\in[1:T]}|\mathcal{P}_t| = T^2 Q$, and with (36), we obtain $|\mathcal{P}_t| = TQ$. Furthermore, from (33), $\mathcal{I}_r = [1\!:\!L]$ for $r \in [1\!:\!T]$. We choose $\boldsymbol{s}_{r,t} = \boldsymbol{0}$ for $r \neq t$, and we choose $[\boldsymbol{Z}_{r,t}]_{\mathcal{P}_r}$ such that $\big[\big(\boldsymbol{Z}_{r,1} \cdots \boldsymbol{Z}_{r,T}\big)\big]_{\mathcal{P}_r}$ is nonsingular. We have $[\boldsymbol{A}_{r,t}]_{\mathcal{P}_t} = \boldsymbol{0}$ (cf. (21), noting that $\mathcal{P}_t \cap \mathcal{D}_t = \emptyset$). Hence, we can use Lemma 6 with $\boldsymbol{M} \triangleq \det(\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s},\boldsymbol{x}_\mathcal{D}))$ given by (22) and $[\boldsymbol{M}]^{\mathcal{J}}_{\mathcal{I}} = \operatorname{diag}\big(\big[\big(\boldsymbol{Z}_{1,1} \cdots \boldsymbol{Z}_{1,T}\big)\big]_{\mathcal{P}_1}, \ldots, \big[\big(\boldsymbol{Z}_{T,1} \cdots \boldsymbol{Z}_{T,T}\big)\big]_{\mathcal{P}_T}\big)$. It thus remains to show that the determinant of the matrix $[\boldsymbol{M}]^{[1:N]\backslash\mathcal{J}}_{[1:N]\backslash\mathcal{I}}$ corresponding to

$$\begin{pmatrix} [\boldsymbol{A}_{1,1}]_{\mathcal{D}_1} & & \\ & \ddots & \\ & & [\boldsymbol{A}_{T,T}]_{\mathcal{D}_T} \end{pmatrix} \quad (37)$$

is nonzero. Because of (21), this matrix is a diagonal matrix and can be chosen to have nonzero elements by choosing $[\boldsymbol{Z}_{t,t}]_{\mathcal{D}_t}$ and $\boldsymbol{s}_{t,t}$ such that $[\boldsymbol{Z}_{t,t}]_{\{i\}}\boldsymbol{s}_{t,t} \neq 0$ for all $i \in \mathcal{D}_t$. Thus, the matrix in (37) is a diagonal matrix with nonzero entries and hence its determinant is nonzero.

Inductive step: We have to show that we can find $\boldsymbol{Z}_{R,t}$ and $\boldsymbol{s}_{R,t}$ for $t \in [1:T]$ such that the determinant of the matrix $\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s},\boldsymbol{x}_\mathcal{D})$ in (22) is nonzero assuming that it is nonzero for the R–1 setting. Let $\mathcal{G}$, $\mathcal{G}_t$, and $\mathcal{L}_t$ be as in Lemma 5 and let $g_t \in \mathcal{G}_t \cap \mathcal{P}_t$ ($\neq \emptyset$ due to Lemma 5). Set $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}\backslash\mathcal{G}_t} = \boldsymbol{0}$. Furthermore, let $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}_t}$ be nonsingular for all $t \in [1:T]$. It easily follows that $\big([\boldsymbol{Z}_{R,1}]_\mathcal{G} \cdots [\boldsymbol{Z}_{R,T}]_\mathcal{G}\big)$ is nonsingular. Next, we choose $\boldsymbol{s}_{R,t}$ such that it is orthogonal to the rows of $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}_t\backslash\{g_t\}}$ and satisfies $[\boldsymbol{Z}_{R,t}]_{\{g_t\}}\boldsymbol{s}_{R,t} \neq 0$. With (21) and $g_t \in \mathcal{P}_t$, we then obtain $[\boldsymbol{A}_{R,t}]_\mathcal{G} = \boldsymbol{0}$, $t \in [1:T]$. Hence, according to Lemma 6 with $\boldsymbol{M}$ given by (22) and $[\boldsymbol{M}]^{\mathcal{J}}_{\mathcal{I}} = \big([\boldsymbol{Z}_{R,1}]_\mathcal{G} \cdots [\boldsymbol{Z}_{R,T}]_\mathcal{G}\big)$, the determinant of $\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s},\boldsymbol{x}_\mathcal{D})$ in (22) is nonzero if and only if the determinant of the following matrix is nonzero:

$$\begin{pmatrix} & & & [\boldsymbol{A}_{1,1}]_{\mathcal{I}_1} & \cdots & [\boldsymbol{A}_{1,T}]_{\mathcal{I}_1} \\ \hat{\boldsymbol{\Xi}}_1 & \cdots & \hat{\boldsymbol{\Xi}}_T & \vdots & & \vdots \\ & & & [\boldsymbol{A}_{R-1,1}]_{\mathcal{I}_{R-1}} & \cdots & [\boldsymbol{A}_{R-1,T}]_{\mathcal{I}_{R-1}} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & [\boldsymbol{A}_{R,1}]_{\bigcup_{t\in[1:T]}\mathcal{L}_t} & \cdots & [\boldsymbol{A}_{R,T}]_{\bigcup_{t\in[1:T]}\mathcal{L}_t} \end{pmatrix},$$

where

$$\hat{\boldsymbol{\Xi}}_t \triangleq \begin{pmatrix} [\boldsymbol{Z}_{1,t}]_{\mathcal{I}_1} & & \\ & \ddots & \\ & & [\boldsymbol{Z}_{R-1,t}]_{\mathcal{I}_{R-1}} \end{pmatrix}.$$

By choosing the remaining rows of $\boldsymbol{Z}_{R,t}$ appropriately, we obtain $[\boldsymbol{A}_{R,t}]_{(\bigcup_{t'\in[1:T]}\mathcal{L}_{t'})\backslash\mathcal{L}_t} = \boldsymbol{0}$ and $\det\big([\boldsymbol{A}_{R,t}]^{\mathcal{L}_t}_{\mathcal{L}_t}\big) \neq 0$. By Lemma 6, it can then be easily seen that the determinant of $\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s},\boldsymbol{x}_\mathcal{D})$ in (22) is nonzero if and only if the determinant of (22) for $R-1$ is nonzero, which is true by the induction hypothesis.

Appendix B: Proof of Lemma 3

First, we state the version of the change-of-variables theorem [10, Theorem 3.2.5] that we will use.

Lemma 7: Let $\psi\colon \mathbb{C}^n \to \mathbb{C}^n$ be a differentiable mapping with Jacobian matrix $\boldsymbol{J}_\psi$. Then for any measurable, nonneg-

ative, real-valued function g on $\mathbb{C}^n$ and any measurable set $\mathcal{S} \subseteq \mathbb{C}^n$, we have

$$\int_{\mathcal{S}} g(\psi(\boldsymbol{u}))\,|\boldsymbol{J}_{\psi}(\boldsymbol{u})|^2\,d\boldsymbol{u} \;=\; \int_{\mathbb{C}^n} g(\boldsymbol{v})\,\mathrm{Nr}(\psi\,|\,\mathcal{S},\boldsymbol{v})\,d\boldsymbol{v}\,,$$

where $\mathrm{Nr}(\psi\,|\,\mathcal{S},\boldsymbol{v})$ denotes the number of points $\boldsymbol{u} \in \mathcal{S}$ such that $\psi(\boldsymbol{u}) = \boldsymbol{v}$. (Note, in particular, that $\mathrm{Nr}(\psi\,|\,\mathcal{S},\boldsymbol{v}) = 0$ if there is no $\boldsymbol{u} \in \mathcal{S}$ such that $\psi(\boldsymbol{u}) = \boldsymbol{v}$.)

We will also make use of the following lemma to obtain one-to-one mappings with maximal support.

Lemma 8: For any Lebesgue measurable set $\mathcal{A} \subseteq \mathbb{C}^n$ and any mapping $\psi\colon \mathbb{C}^n \to \mathbb{C}^n$ such that $|\psi^{-1}(\{\boldsymbol{v}\}) \cap \mathcal{A}| \le m < \infty$ for all $\boldsymbol{v} \in \mathbb{C}^n$, there exists a Lebesgue measurable set $\mathcal{B} \subseteq \mathcal{A}$ such that $\psi\big|_{\mathcal{B}}$ is one-to-one and $\psi(\mathcal{B}) = \psi(\mathcal{A})$. Furthermore, $|\psi^{-1}(\{\boldsymbol{v}\}) \cap (\mathcal{A}\backslash\mathcal{B})| \le m-1 < \infty$ for all $\boldsymbol{v} \in \mathbb{C}^n$.

Proof: Let $\mathfrak{M}$ denote the set of all measurable subsets $\mathcal{V} \subseteq \mathcal{A}$ such that $\psi\big|_{\mathcal{V}}$ is one-to-one. We have the natural partial order of inclusion on $\mathfrak{M}$. For any chain (i.e., totally ordered set) $\mathfrak{C}$ of sets in $\mathfrak{M}$, the union of all sets in $\mathfrak{C}$ is an upper bound for all sets in $\mathfrak{C}$ (i.e., for any $\mathcal{A}_0 \in \mathfrak{C}$ we have $\mathcal{A}_0 \subseteq \bigcup_{\mathcal{C}\in\mathfrak{C}} \mathcal{C}$) and belongs to $\mathfrak{M}$. Thus, by Zorn's lemma, there exists at least one maximal element in $\mathfrak{M}$. Let $\mathcal{B}$ be a maximal element in $\mathfrak{M}$. If there exists a $\boldsymbol{v} \in \psi(\mathcal{A})\backslash\psi(\mathcal{B})$, we can add one point $\boldsymbol{u} \in \psi^{-1}(\{\boldsymbol{v}\})$ to $\mathcal{B}$ and $\mathcal{B} \cup \{\boldsymbol{u}\}$ belongs to $\mathfrak{M}$ with $\mathcal{B} \subsetneq \mathcal{B} \cup \{\boldsymbol{u}\}$, which is a contradiction to the maximality of $\mathcal{B}$. Hence, $\psi(\mathcal{B}) = \psi(\mathcal{A})$. Furthermore, since $\mathcal{B} \in \mathfrak{M}$ the set $\mathcal{B}$ is measureable and $\psi\big|_{\mathcal{B}}$ is one-to-one. Finally, for each $\boldsymbol{v} \in \psi(\mathcal{A})$, there exists a $\boldsymbol{u} \in \mathcal{B}$ such that $\psi(\boldsymbol{u}) = \boldsymbol{v}$. Thus, $|\psi^{-1}(\{\boldsymbol{v}\}) \cap (\mathcal{A}\backslash\mathcal{B})| \le |\psi^{-1}(\{\boldsymbol{v}\}) \cap \mathcal{A}| - 1 \le m-1$. ■

For $\mathcal{M}$ and m as defined in Lemma 3, we now partition the set $\mathcal{M}$ into subsets $\mathcal{V}_i$ with $i \in [1:m]$ such that each $\kappa_i \triangleq \kappa\big|_{\mathcal{V}_i}$ is one-to-one and $\mathbb{C}^n\backslash\bigcup_{i\in[1:m]}\mathcal{V}_i$ has Lebesgue measure zero. The existence of such sets can be shown by using Lemma 8 repeatedly. Next, we define the set $\mathcal{U}$ used in Lemma 3. Let

$$\tilde{\mathcal{U}} \;\triangleq\; \bigg\{\boldsymbol{u} \in \mathcal{M} : \frac{f_{\mathbf{u}}(\boldsymbol{u})}{|\boldsymbol{J}_{\kappa}(\boldsymbol{u})|^2} \ge \frac{f_{\mathbf{u}}(\tilde{\boldsymbol{u}})}{|\boldsymbol{J}_{\kappa}(\tilde{\boldsymbol{u}})|^2} \quad \forall\, \tilde{\boldsymbol{u}} \in \kappa^{-1}(\kappa(\{\boldsymbol{u}\})) \cap \mathcal{M}\bigg\}. \tag{38}$$

Note that $\kappa(\tilde{\mathcal{U}}) = \kappa(\mathcal{M})$. The set $\tilde{\mathcal{U}}$ is measurable since it is the preimage of $\{1\}$ under the measurable function[5]

$$g(\boldsymbol{u}) \;\triangleq\; \frac{\frac{f_{\mathbf{u}}(\boldsymbol{u})}{|\boldsymbol{J}_{\kappa}(\boldsymbol{u})|^2}}{\max_{i\in\mathcal{F}(\boldsymbol{u})} \frac{f_{\mathbf{u}}(\kappa_i^{-1}(\kappa(\boldsymbol{u})))}{|\boldsymbol{J}_{\kappa}(\kappa_i^{-1}(\kappa(\boldsymbol{u})))|^2}}\,,$$

where $\mathcal{F}(\boldsymbol{u}) \triangleq \{i \in [1:m] : \kappa(\boldsymbol{u}) \in \kappa_i(\mathcal{V}_i)\}$. By Lemma 8 with $\psi = \kappa$ and $\mathcal{A} = \tilde{\mathcal{U}}$, there exists a set $\mathcal{U} \subseteq \tilde{\mathcal{U}}$ such that $\kappa\big|_{\mathcal{U}}$ is one-to-one and $\kappa(\mathcal{U}) = \kappa(\tilde{\mathcal{U}}) = \kappa(\mathcal{M})$. Applying Lemma 7 with $g(\boldsymbol{v}) = -f_{\mathbf{v}}(\boldsymbol{v})\log(f_{\mathbf{v}}(\boldsymbol{v}))$, $\psi = \kappa$, and $\mathcal{S} = \mathcal{U}$ yields ($f_{\mathbf{v}}$ denotes the density of $\mathbf{v} = \kappa(\mathbf{u})$)

$$\begin{aligned} h(\mathbf{v}) \;&=\; -\int_{\mathbb{C}^n} f_{\mathbf{v}}(\boldsymbol{v})\log(f_{\mathbf{v}}(\boldsymbol{v}))\,d\boldsymbol{v} \\ &\overset{(a)}{=} -\int_{\mathbb{C}^n} f_{\mathbf{v}}(\boldsymbol{v})\log(f_{\mathbf{v}}(\boldsymbol{v}))\,\mathrm{Nr}(\kappa|\mathcal{U},\boldsymbol{v})\,d\boldsymbol{v} \\ &= -\int_{\mathcal{U}} f_{\mathbf{v}}(\kappa(\boldsymbol{u}))\log(f_{\mathbf{v}}(\kappa(\boldsymbol{u})))\,|\boldsymbol{J}_{\kappa}(\boldsymbol{u})|^2\,d\boldsymbol{u}\,. \end{aligned} \tag{39}$$

Here, (a) holds because $\boldsymbol{v}$ is supported (up to a set of measure zero) on $\kappa(\mathcal{U})$; note also that $\mathrm{Nr}(\kappa|\mathcal{U},\boldsymbol{v})$ is 1 for $\boldsymbol{v} = \kappa(\boldsymbol{u})$ and 0 else. The next step is to establish a relation between the densities $f_{\mathbf{v}}(\kappa(\boldsymbol{u}))$ and $f_{\mathbf{u}}(\boldsymbol{u})$ for $\boldsymbol{u} \in \mathcal{U}$. Let $\mathcal{U}' \subseteq \mathcal{U}$ be any measurable subset of $\mathcal{U}$. We have

$$\begin{aligned} \int_{\mathcal{U}'} f_{\mathbf{v}}(\kappa(\boldsymbol{u}))\,|\boldsymbol{J}_{\kappa}(\boldsymbol{u})|^2\,d\boldsymbol{u} \;&=\; \int_{\kappa(\mathcal{U}')} f_{\mathbf{v}}(\boldsymbol{v})\,d\boldsymbol{v} \\ &= \Pr\{\boldsymbol{v} \in \kappa(\mathcal{U}')\} \\ &= \Pr\{\boldsymbol{u} \in \kappa^{-1}(\kappa(\mathcal{U}'))\} \\ &= \int_{\kappa^{-1}(\kappa(\mathcal{U}'))} f_{\mathbf{u}}(\boldsymbol{u})\,d\boldsymbol{u}\,. \end{aligned} \tag{40}$$

Since $\kappa_i = \kappa\big|_{\mathcal{V}_i}$, we have

$$\bigcup_{i\in[1:m]} \kappa_i^{-1}(\kappa(\mathcal{U}')) \;=\; \bigcup_{i\in[1:m]} \big(\kappa^{-1}(\kappa(\mathcal{U}')) \cap \mathcal{V}_i\big)\,,$$

and since $\mathbb{C}^n\backslash\bigcup_{i\in[1:m]}\mathcal{V}_i$ has Lebesgue measure zero, the set $\bigcup_{i\in[1:m]}\kappa_i^{-1}(\kappa(\mathcal{U}'))$ is equal to $\kappa^{-1}(\kappa(\mathcal{U}'))$ up to a set of Lebesgue measure zero. Thus,

$$\int_{\kappa^{-1}(\kappa(\mathcal{U}'))} f_{\mathbf{u}}(\boldsymbol{u})\,d\boldsymbol{u} \;=\; \sum_{i\in[1:m]} \int_{\kappa_i^{-1}(\kappa(\mathcal{U}'))} f_{\mathbf{u}}(\boldsymbol{u})\,d\boldsymbol{u} \tag{41}$$

(note that $\kappa_i^{-1}(\kappa(\mathcal{U}')) \subseteq \mathcal{V}_i$ and the $\mathcal{V}_i$ are disjoint). Using for an arbitrary $i \in [1:m]$ Lemma 7 with $\psi = \kappa_i^{-1}$ and $\mathcal{S} = \kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}')))$, and using the inverse function theorem, we obtain (note that $\boldsymbol{J}_{\kappa} = \boldsymbol{J}_{\kappa_i}$ on $\mathcal{V}_i$ because $\kappa_i = \kappa\big|_{\mathcal{V}_i}$)

$$\int_{\kappa_i^{-1}(\kappa(\mathcal{U}'))} f_{\mathbf{u}}(\boldsymbol{u})\,d\boldsymbol{u} \;=\; \int_{\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}')))} \frac{f_{\mathbf{u}}(\kappa_i^{-1}(\boldsymbol{v}))}{|\boldsymbol{J}_{\kappa}(\kappa_i^{-1}(\boldsymbol{v}))|^2}\,d\boldsymbol{v}\,. \tag{42}$$

Another application of Lemma 7 with $\psi = \tilde{\kappa} \triangleq \kappa\big|_{\mathcal{U}}$ and $\mathcal{S} = \tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}'))))$ then gives

$$\begin{aligned} &\int_{\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}')))} \frac{f_{\mathbf{u}}(\kappa_i^{-1}(\boldsymbol{v}))}{|\boldsymbol{J}_{\kappa}(\kappa_i^{-1}(\boldsymbol{v}))|^2}\,d\boldsymbol{v} \\ &\quad= \int_{\tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}'))))} \frac{f_{\mathbf{u}}(\kappa_i^{-1}(\tilde{\kappa}(\tilde{\boldsymbol{u}})))\,|\boldsymbol{J}_{\kappa}(\tilde{\boldsymbol{u}})|^2}{|\boldsymbol{J}_{\kappa}(\kappa_i^{-1}(\tilde{\kappa}(\tilde{\boldsymbol{u}})))|^2}\,d\tilde{\boldsymbol{u}}\,. \end{aligned} \tag{43}$$

We can upper-bound (43) by

$$\begin{aligned} &\int_{\tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}'))))} \frac{f_{\mathbf{u}}(\kappa_i^{-1}(\tilde{\kappa}(\tilde{\boldsymbol{u}})))\,|\boldsymbol{J}_{\kappa}(\tilde{\boldsymbol{u}})|^2}{|\boldsymbol{J}_{\kappa}(\kappa_i^{-1}(\tilde{\kappa}(\tilde{\boldsymbol{u}})))|^2}\,d\tilde{\boldsymbol{u}} \\ &\quad\overset{(a)}{\le} \int_{\tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}'))))} \frac{f_{\mathbf{u}}(\tilde{\boldsymbol{u}})\,|\boldsymbol{J}_{\kappa}(\tilde{\boldsymbol{u}})|^2}{|\boldsymbol{J}_{\kappa}(\tilde{\boldsymbol{u}})|^2}\,d\tilde{\boldsymbol{u}} \end{aligned}$$

[5] The function g is measurable by the following argument: κ_i^{-1} is continuous by the inverse function theorem. Hence, for all $\boldsymbol{u}$ with equal $\mathcal{F}(\boldsymbol{u})$, the denominator in the definition of g is just the maximum over a finite set of continuous functions and thus measureable. Since there are only a finite number of possible realizations of $\mathcal{F}(\boldsymbol{u})$, we can partition the domain of g into a finite number of sets where g is measureable. Therefore, g is measureable.

$$
\begin{aligned}
&= \int_{\tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}'))))} f_{\mathsf{u}}(\tilde{\boldsymbol{u}})\, d\tilde{\boldsymbol{u}} \\
&\overset{(b)}{\leq} \int_{\mathcal{U}'} f_{\mathsf{u}}(\tilde{\boldsymbol{u}})\, d\tilde{\boldsymbol{u}}\,,
\end{aligned} \tag{44}
$$

where in (a) we used the fact that $\tilde{\boldsymbol{u}} \in \tilde{\mathcal{U}}$ (we have $\tilde{\boldsymbol{u}} \in \tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}')))) = \tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\tilde{\kappa}(\mathcal{U}')))) = (\tilde{\kappa}^{-1} \circ \kappa_i)((\tilde{\kappa}^{-1} \circ \kappa_i)^{-1}(\mathcal{U}')) \subseteq \mathcal{U}' \subseteq \mathcal{U} \subseteq \tilde{\mathcal{U}}$) and the inequality in (38), and in (b) we used $\tilde{\kappa}^{-1}(\kappa_i(\kappa_i^{-1}(\kappa(\mathcal{U}')))) \subseteq \mathcal{U}'$. Note that the upper bound (44) does not depend on $i \in [1:m]$. Hence, (40)–(44) yield

$$
\int_{\mathcal{U}'} f_{\mathsf{v}}(\kappa(\boldsymbol{u}))\, |\boldsymbol{J}_\kappa(\boldsymbol{u})|^2\, d\boldsymbol{u} \,\leq\, m \int_{\mathcal{U}'} f_{\mathsf{u}}(\boldsymbol{u})\, d\boldsymbol{u}\,,
$$

for an arbitrary measurable set $\mathcal{U}' \subseteq \mathcal{U}$. Thus,

$$
f_{\mathsf{v}}(\kappa(\boldsymbol{u}))\, |\boldsymbol{J}_\kappa(\boldsymbol{u})|^2 \,\leq\, m\, f_{\mathsf{u}}(\boldsymbol{u}) \quad \text{a.e. on } \mathcal{U}\,.
$$

Inserting this into (39) leads to

$$
\begin{aligned}
h(\mathsf{v}) &\geq -\int_{\mathcal{U}} m\, f_{\mathsf{u}}(\boldsymbol{u}) \log\left(\frac{m f_{\mathsf{u}}(\boldsymbol{u})}{|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2}\right) d\boldsymbol{u} \\
&\geq -\, m \log(m) - m \int_{\mathcal{U}} f_{\mathsf{u}}(\boldsymbol{u}) \log(f_{\mathsf{u}}(\boldsymbol{u}))\, d\boldsymbol{u} \\
&\quad + m \int_{\mathcal{U}} f_{\mathsf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\, d\boldsymbol{u}\,.
\end{aligned}
$$

Appendix C: Proof of Lemma 4

Since f is not identically zero, there is a $\boldsymbol{\xi}_0 \in \mathbb{C}^N$ such that $f(\boldsymbol{\xi}_0) \neq 0$. Then $g(\boldsymbol{\xi}) \triangleq f(\boldsymbol{\xi} + \boldsymbol{\xi}_0)$ is an analytic function that is nonzero at $\boldsymbol{\xi} = \mathbf{0}$. By changing variables $\boldsymbol{\xi} \mapsto \boldsymbol{\xi} + \boldsymbol{\xi}_0$, we obtain for I_1 in (32)

$$
I_1 \,=\, \int_{\mathbb{C}^N} \exp(-\|\boldsymbol{\xi} + \boldsymbol{\xi}_0\|^2) \log(|g(\boldsymbol{\xi})|)\, d\boldsymbol{\xi}\,.
$$

Noting that

$$
\begin{aligned}
\|\boldsymbol{\xi} + \boldsymbol{\xi}_0\|^2 &\leq \|\boldsymbol{\xi}\|^2 + 2\|\boldsymbol{\xi}\|\|\boldsymbol{\xi}_0\| + \|\boldsymbol{\xi}_0\|^2 \\
&\leq \|\boldsymbol{\xi}\|^2 + 2\max\{\|\boldsymbol{\xi}\|^2, \|\boldsymbol{\xi}_0\|^2\} + \|\boldsymbol{\xi}_0\|^2 \\
&\leq 3\|\boldsymbol{\xi}\|^2 + 3\|\boldsymbol{\xi}_0\|^2\,,
\end{aligned}
$$

we can lower bound I_1 by

$$
I_1 \,\geq\, c \int_{\mathbb{C}^N} \exp(-3\|\boldsymbol{\xi}\|^2) \log(|g(\boldsymbol{\xi})|)\, d\boldsymbol{\xi} \,\triangleq\, I_2\,, \tag{45}
$$

with $c \triangleq \exp(-3\|\boldsymbol{\xi}_0\|^2)$. Using the mapping $\varphi\colon \mathbb{R}^{2N} \to \mathbb{C}^N$; $\boldsymbol{x} \mapsto \big(\boldsymbol{x}_{[1:N]} + i\boldsymbol{x}_{[N+1:2N]}\big)$, we can write I_2 in (45) as

$$
I_2 \,=\, c \int_{\mathbb{R}^{2N}} \exp(-3\|\boldsymbol{x}\|^2)\, u(\boldsymbol{x})\, d\boldsymbol{x}\,, \tag{46}
$$

with $u(\boldsymbol{x}) \triangleq \log(|g(\varphi(\boldsymbol{x}))|)$. Since $g(\mathbf{0}) \neq 0$, we have $u(\mathbf{0}) > -\infty$. By [12, Example 2.6.1.3], $u(\boldsymbol{x})$ is a subharmonic function. A useful property of subharmonic functions is stated in the following lemma (see [12, Theorem 2.6.2.1]).

Lemma 9: Let u be a subharmonic function on $\mathcal{W} \subseteq \mathbb{R}^{2N}$, and let $\boldsymbol{x} \in \mathbb{R}^{2N}$. If $\mathcal{B}_{\boldsymbol{x},r} \subseteq \mathcal{W}$ for some $r > 0$, with $\mathcal{B}_{\boldsymbol{x},r} \triangleq \{\boldsymbol{v} \in \mathbb{R}^{2N} : \|\boldsymbol{v} - \boldsymbol{x}\| \leq r\}$, then

$$
u(\boldsymbol{x}) \,\leq\, \frac{1}{\sigma_{2N}\, r^{2N-1}} \int_{\mathcal{S}_{\boldsymbol{x},r}} u(\boldsymbol{y})\, ds(\boldsymbol{y})\,,
$$

where $\mathcal{S}_{\boldsymbol{x},r} \triangleq \{\boldsymbol{y} \in \mathbb{R}^{2N} : \|\boldsymbol{y} - \boldsymbol{x}\| = r\}$, σ_{2N} is the area of the unit sphere in $\mathbb{R}^{2N}$, and ds denotes integration with respect to the $(2N-1)$-dimensional Hausdorff measure (cf. [10, Subsection 2.10.2]).

Using a well-known measure-theoretic result [10, Theorem 3.2.12], we obtain

$$
\begin{aligned}
&\int_{\mathbb{R}^{2N}} \exp(-3\|\boldsymbol{x}\|^2)\, u(\boldsymbol{x})\, d\boldsymbol{x} \\
&\qquad = \int_{(0,\infty)} \int_{\mathcal{S}_{\mathbf{0},r}} \exp(-3r^2)\, u(\boldsymbol{y})\, ds(\boldsymbol{y})\, dr\,.
\end{aligned} \tag{47}
$$

We thus have

$$
\begin{aligned}
I_2 &\overset{(a)}{=} c \int_{(0,\infty)} \int_{\mathcal{S}_{\mathbf{0},r}} \exp(-3r^2)\, u(\boldsymbol{y})\, ds(\boldsymbol{y})\, dr \\
&\overset{(b)}{\geq} c\, \sigma_{2N}\, u(\mathbf{0}) \int_{(0,\infty)} \exp(-3r^2)\, r^{2N-1}\, dr \\
&\overset{(c)}{>} -\infty\,,
\end{aligned}
$$

where (a) follows by using (47) in (46), (b) is due to Lemma 9, and (c) holds because $u(\mathbf{0}) > -\infty$. With (45), it then follows that $I_1 > -\infty$.

Acknowledgment

We wish to thank Dr. Shaowei Lin for kindly pointing us to the weak version of Bézout's theorem.

References

[1] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Th.*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[2] V. I. Morgenshtern, G. Durisi, and H. Bölcskei, "The SIMO pre-log can be larger than the SISO pre-log," in *Proc. IEEE Int. Symp. Inf. Th. (ISIT 2010)*, Austin, TX, June 2010, pp. 320–324.

[3] E. Riegler, V. I. Morgenshtern, G. Durisi, S. Lin, B. Sturmfels, and H. Bölcskei, "Noncoherent SIMO pre-log via resolution of singularities," in *Proc. IEEE Int. Symp. Inf. Th. (ISIT 2011)*, St. Petersburg, Russia, Aug. 2011, pp. 2020–2024.

[4] W. Yang, G. Durisi, V. I. Morgenshtern, and E. Riegler, "Capacity pre-log of SIMO correlated block-fading channels," in *Proc. IEEE Int. Symp. Wireless Comm. Syst. (ISWCS 2011)*, Aachen, Germany, Nov. 2011, pp. 869–873.

[5] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Th.*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[6] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Th.*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[7] A. R. P. van den Essen, *Polynomial Automorphisms and the Jacobian Conjecture*. Basel, Switzerland: Birkhäuser, 2000.

[8] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: Wiley, 1968.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley, 2006.

[10] H. Federer, *Geometric Measure Theory*. New York, NY: Springer, 1969.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1985.

[12] V. Azarin, *Growth Theory of Subharmonic Functions*. Basel, Switzerland: Birkhäuser, 2009.